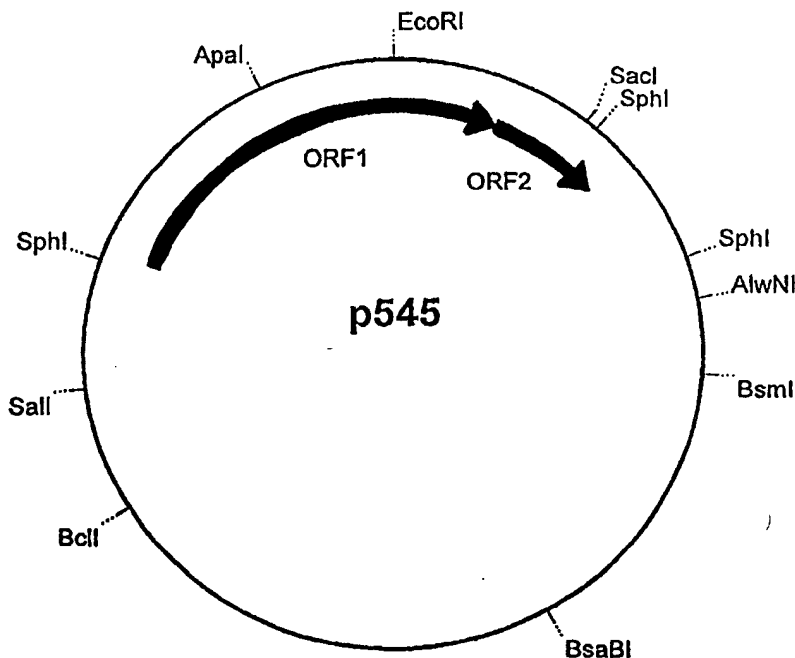




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N	A2	(11) International Publication Number: WO 99/67356 (43) International Publication Date: 29 December 1999 (29.12.99)
<p>(21) International Application Number: PCT/EP99/04416</p> <p>(22) International Filing Date: 25 June 1999 (25.06.99)</p> <p>(30) Priority Data: 98305033.7 25 June 1998 (25.06.98) EP</p> <p>(71) Applicant (for all designated States except US): GIST-BROCADES B.V. [NL/NL]; Wateringseweg 1, P.O. Box 1, NL-2600 MA Delft (NL).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): POUWELS, Pieter, Hendrik [NL/NL]; Delftweg 14, NL-2289 AJ Rijswijk (NL). VAN LUIJK, Nicole [NL/NL]; Wageningsberg 34, NL-3524 LR Utrecht (NL). JORE, Johannes, Petrus, Maria [NL/NL]; Waterruit 9, NL-2804 PA Gouda (NL). LUITEN, Rudolf, Gijsbertus, Marie [NL/NL]; Akkerhoornbloem 26, NL-2317 KR Leiden (NL).</p> <p>(74) Agents: WRIGHT, Simon, Mark et al.; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5LX (GB).</p>	<p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i> <i>With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.</i></p>	

(54) Title: PROPIONIBACTERIUM VECTOR



(57) Abstract

An endogenous plasmid of Propionibacterium is described, isolated from Propionibacteria freudenreichii LMG 16545 (deposited as CBS 101022), and its sequence provided. This plasmid can be used to transform Propionibacteria to express homologous or heterologous proteins, in the production of recombinant proteins or products of enzymes, for example vitamin B₁₂.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PROPIONIBACTERIUM VECTOR

This invention relates to an endogenous plasmid of *Propionibacterium*, vectors derived from it and the use of these vectors to express (heterologous) proteins in bacteria, especially *Propionibacteria*. In particular transformed bacteria can be used either to
5 produce, by fermentation, vitamin B₁₂ or in cheese making.

Introduction

Propionibacteria are Gram-positive bacteria capable of producing various useful compounds in a variety of industrial processes. For example, several *Propionibacterium* species are known to produce vitamin B₁₂ (cobalamin) in large scale fermentation
10 processes. Other species are used in dairy applications such as cheese manufacturing where they contribute, and in many cases even are mainly responsible, for the specific flavour and texture of the cheese. Many *Propionibacterium* species are considered safe for inclusion, as live organisms, into food and animal feed.

To be able to fully exploit the biotechnological potential of *Propionibacterium*,
15 efficient and flexible genetic engineering techniques are required. Such techniques rely on the availability of a suitable plasmid to express a protein from a heterologous gene in *Propionibacterium*.

EP-A-0400931 (Nippon Oil) refers to an endogenous plasmid (pTY-1) from *Propionibacterium pentosaceum* (ATCC 4875) but does not describe its sequence or
20 exemplify how it may be used to express a heterologous gene.

JP 08-56673 refers to the plasmid pTY-1 for producing vitamin B₁₂ but does not provide any evidence that the plasmid remains as a freely replicating extrachromosomal element nor that the plasmid is stable inside the transformed cells.

The invention therefore seeks to provide vectors that are more efficient than those in
25 the prior art, and can remain extrachromosomal and/or are stable. In particular the invention aims to provide an efficient vector for the cloning or expression of *Propionibacterium* or foreign genomic fragments or genes into a (*Propionibacterium*) host strain. This may enable host specific restriction enzymes to be circumvented and thereby avoid the host treating the plasmid as a foreign polynucleotide.

Summary of the invention

Accordingly, the present invention in a first aspect provides a polynucleotide comprising a sequence capable of hybridising selectively to

- 5 (a) SEQ ID NO: 1 or the complement thereof;
- (b) a sequence from the 3.6 kb plasmid of *Propionibacterium freudenreichii* CBS 101022;
- (c) a sequence from the 3.6 kb plasmid of *Propionibacterium freudenreichii* CBS 101023; or
- 10 (d) a sequence that encodes a polypeptide of the invention, such as (at least part of) the amino acid sequence of SEQ ID NO: 2 or SEQ ID No: 3, or the complement thereof.

SEQ ID NO: 1 sets out the DNA sequence of the endogenous plasmid of *Propionibacterium* LMG 16545 which the inventors have discovered. The first coding
15 sequence runs from nucleotide 273 to nucleotide 1184 and the predicted amino acid sequence of this coding sequence is shown in SEQ ID NO: 2. The second coding sequence runs from nucleotides 1181 to 1483 and the predicted amino acid sequence of this coding sequence is shown in SEQ ID No: 3.

The inventors screened a large collection of *Propionibacterium* isolates and identified
20 two strains both harboring cryptic plasmids with a size of 3.6 kb. One of the strains is *Propionibacterium freudenreichii* LMG 16545 which was deposited at Centraalbureau voor Schimmelcultures (CBS), Oosterstraat 1, Postbus 273, NL-3740 AG Baarn, Netherlands, in the name of Gist-brocades B.V. of Wateringseweg 1, P.O. Box 1, 2600 MA Delft, The Netherlands, on 19 June 1998 under the terms of the Budapest Treaty and was
25 given accession number CBS 101022. The other strain is *Propionibacterium freudenreichii* LMG 16546 which was also deposited by the same depositor on 19 June 1998 under the terms of the Budapest Treaty at CBS and was given accession number CBS 101023.

Through full characterization and computer assisted analysis of the nucleotide
30 sequence of LMG 16545 the inventors have been able to identify insertion sites for foreign DNA fragments. These sites have allowed construction of plasmids that are still capable of autonomous replication in *Propionibacterium*.

Surprisingly it was found that an erythromycin resistance gene from the actinomycete *Saccharopolyspora erythraea* is efficiently expressed in *Propionibacterium* and thus can be used as a selection marker for transformed cells.

Also constructed are bifunctional vectors, stably maintainable and selectable in both *E. coli* and *Propionibacterium*. This allows the use of *E. coli* for vector construction, as well as functional expression of homologous or heterologous genes in *Propionibacterium*. Vector construction using *E. coli* is comparatively easy and can be done quickly.

The polynucleotide of the invention may be autonomously replicating or extrachromosomal, for example in a bacterium such as a *Propionibacterium*.

Hence another aspect the invention provides a vector which comprises a polynucleotide of the invention.

The invention also provides a process for the preparation of a polypeptide, the process comprising cultivating a host cell transformed or transfected with a vector of the invention under conditions to provide for expression of the polypeptide.

The invention additionally provides a polypeptide which comprises the sequence set out in SEQ ID NO: 2 or 3 or a sequence substantially homologous to that sequence, or a fragment of either sequence, or a protein encoded by a polynucleotide of the invention.

Detailed Description of the Invention

Polynucleotides

A polynucleotide of the invention may be capable of hybridising selectively with the sequence of SEQ ID NO: 1, or a portion of SEQ ID No: 1, or to the sequence complementary to that sequence or portion of the sequence. The polynucleotide of the invention may be capable of hybridising selectively to the sequence of the 3.6 kb plasmid of *P. freudenreichii* CBS 101022 or CBS 101023, or to a portion of the sequence of either plasmid. Typically, a polynucleotide of the invention is a contiguous sequence of nucleotides which is capable of selectively hybridizing to the sequence of SEQ ID. No: 1 or of either 3.6 kb plasmid, or portion of any of these sequences, or to the complement of these sequences or portion of any of these sequences.

A polynucleotide of the invention and the sequence of SEQ ID NO: 1, or either

of the 3.6 kb plasmids, or a sequence encoding a polypeptide, or a portion of these sequences, can hybridize at a level significantly above background. Background hybridization may occur, for example, because of other polynucleotides present in the preparation. The signal level generated by the interaction between a polynucleotide of the invention and the sequence of SEQ ID NO: 1 or of either 3.6 kb plasmid, or portion of these sequences, is typically at least 10 fold, preferably at least 100 fold, as intense as interactions between other polynucleotides and the coding sequence of SEQ ID NO: 1 or of either 3.6 kb plasmid, or a sequence encoding the polypeptide, or portion of these sequences. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P . Selective hybridisation is typically achieved using conditions of medium stringency (for example 0.3M sodium chloride and 0.03M sodium citrate at about 50°C) to high stringency (same conditions but at about 60°C).

Polynucleotides included in the invention can be generally at least 70%, preferably at least 80 or 90%, more preferably at least 95%, and optimally at least 98% homologous (to the sequences (a) to (d)) over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% or 90% homologous over 25, preferably over 30 nucleotides forms one embodiment of the invention, as does a polynucleotide which is at least 90% or 95% homologous over 40 nucleotides.

The portions referred to above may be the coding sequences of SEQ ID No: 1 or of either 3.6 kb plasmid. Other preferred portions of SEQ ID No: 1 are the replication origin, promoter or regulatory sequences, or sequences capable of effecting or assisting autonomous replication in a host cell, such as a *Propionibacterium*.

It has been found that the portion of the plasmid from the restriction site *SaII* to *AlwNI* appears to be the region that is required for replication of the plasmid. Other parts of the plasmid have been deleted and yet replication does not appear to have been adversely affected. Therefore in the invention sequence (b) and (c) can be the

region delineated by the restriction sites *Sa*II and *Al*uNI. This is approximately 1.7b in length. Alternatively, sequences (b) or (c) can be replaced by the sequence corresponding to nucleotides 1 to 1,800, such as 100 to 1,700, suitably 150 to 1,500, advantageously 200 to 1,300 and optimally 250 to 1,200 of SEQ. ID. No. 1.

5 The proteins (SEQ. ID. Nos. 2 and 3) encoded by ORF1 and ORF2 respectively, are thought to both help the plasmid replicate. The plasmid replicates by the known rolling circle replication method where the original ds DNA plasmid is cut by either of the proteins which assists production of a copy of the outer strand using the inner strand as a template. The copy of the outer ring is removed and the ends joined. The
10 host then replicates a new inner ring using the generated outer ring as a template.

The coding sequences of the invention may be modified by nucleotide substitutions, for example from 1, 2 or 3 to 10, 25, 50 or 100 substitutions. The polynucleotides of sequences (a) to (d) may alternatively or additionally be modified by one or more insertions or deletions and/or by an extension at either or both ends.
15 Degenerate substitutions may be made and/or substitutions may be made which would result in a conservative amino acid substitution when the modified sequence is translated, for example as discussed later with relation to the polypeptides of the invention.

Polynucleotides of the invention may comprise DNA or RNA. They may also be
20 polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to polynucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides
25 described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance *in vivo* activity or lifespan.

Polynucleotides of the invention may be used as a primer, e.g. a PCR (polymerase chain reaction) primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using
30 radioactive or non-radioactive labels, or the polynucleotides may be incorporated or cloned into vectors.

Such primers, probes and other fragments will be at least 15, preferably at least

20, for example at least 25, 30 or 40 nucleotides in length. They will typically be up to 40, 50, 60, 70, 100 or 150 nucleotides in length. Probes and fragments can be longer than 150 nucleotides, for example up to 200, 300, 500, 1,000 or 1,500 nucleotides in length, or even up to a few nucleotides, such as 5 or 10 nucleotides, short of any of the sequences (a) to (d).

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques. The polynucleotides are typically provided in isolated and/or purified form.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR cloning techniques. This will involve making a pair of primers (e.g. of about 15-30 nucleotides) to the region of SEQ ID No: 1 or of either 3.6 kb plasmid which it is desired to clone, bringing the primers into contact with DNA obtained from a *Propionibacterium*, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector. Such techniques may be used to obtain all or part of SEQ ID No: 1 or either 3.6 kb plasmid.

The techniques mentioned herein are well known in the art¹⁰.

Polynucleotides which are not 100% homologous to SEQ ID No: 1 or either 3.6 kb plasmid but fall within the scope of the invention can be obtained in a number of ways.

Homologous polynucleotides of SEQ ID NO: 1 or of either 3.6 kb plasmid may be obtained for example by probing genomic DNA libraries made from a range of *Propionibacteria*, such as *P. freudenreichii*, *P. jensenii*, *P. thoenii*, *P. acidipropionici*, or other strains of bacteria of the class *Actinomycetes*, or other gram positive bacteria,

or those that are G: C rich. All these organisms are suitable sources of homologous or heterologous genes, promoters, enhancers, or host cells, for use in the invention.

Such homologues and fragments thereof in general will be capable of selectively hybridizing to the coding sequence of SEQ ID NO: 1 or its complement
5 or of either 3.6 kb plasmid. Such sequences may be obtained by probing genomic DNA libraries of the *Propionibacterium* with probes comprising all or part of the coding sequence SEQ ID NO: 1 or of either 3.6 kb plasmid under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.3M sodium citrate at from about 50°C to about 60°C).

10 Homologues may also be obtained using degenerate PCR which will use primers designed to target conserved sequences within the homologues. Conserved sequences can be predicted from aligning SEQ ID No: 1 or the sequence of either 3.6 kb plasmid with their homologues. The primers will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning
15 sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of SEQ ID No: 1 or of either 3.6 kb plasmid, or their homologues. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide
20 sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

Methods of measuring polynucleotide homology are well known in the art. For example the UWGCG package provides the BESTFIT programme which can be used
25 to calculate homology, for example used on its default setting⁷. For amino acid homology with regard to polypeptides of the invention which are discussed later, one can employ BLAST (Basic Local Alignment Search Tool¹), which produces alignments of amino acid sequences (and nucleotide sequences if necessary) to determine sequence similarity. BLAST can thus be used to determine exact matches
30 or to identify homologues, and is particularly useful for those matches which do not contain gaps. The BLAST technique uses the algorithm based on the High-scoring Segment Pair (HSP).

The invention includes double stranded polynucleotides comprising a polynucleotide sequence of the invention and its complement.

Polynucleotides (e.g. probes or primers) of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S , enzyme labels, or other
5 protein labels such as biotin. Detection techniques for these labels are known *per se*.

The polynucleotides (labelled or unlabelled) may be used in nucleic acid-based tests for detecting or sequencing another polynucleotide of the invention, in a sample.

Polynucleotides of the invention include variants of the sequence of SEQ ID
10 NO: 1 or of either 3.6 kb plasmid which are capable of autonomously replicating or remaining extrachromosomally in a host cell. Such variants may be stable in a bacterium such as a *Propionibacterium*.

Generally the polynucleotide will comprise the replication origin and/or coding region(s) of SEQ ID No: 1 or of either 3.6 kb plasmid, or homologues of these
15 sequences. A polynucleotide of the invention which is stable in a host cell, such as *Propionibacterium*, or *E. coli* is one which is not lost from the host within five generations, such as fifteen generations, preferably thirty generations. Generally such a polynucleotide would be inherited by both daughter cells every generation.

The polynucleotide may comprise a promoter or an origin of replication (e.g.
20 upstream of any sequences encoding a replication protein).

The polynucleotide of the invention can be transformed or transfected into a bacterium, such as a *Propionibacterium*, or *E. coli*, for example by a suitable method¹¹. It may be present in a bacterium at a copy number of 5 to 500, such as 10 to 100.

25 The polynucleotide may be capable of autonomous replication in a bacterium other than a *Propionibacterium*. Such a bacterium may be *E. coli*, or a gram positive or G:C rich bacterium or one of the class *Actinomycetes*. Such a polynucleotide will generally comprise sequences which enable the polynucleotide to be autonomously replicated in that bacterium. Such sequences can be derived from plasmids which are
30 able to replicate in that bacterium.

A polynucleotide of the invention may be one which has been produced by replication in a *Propionibacterium*. Alternatively it may have been produced by

replication in another bacterium, such as *E. coli*. The polynucleotide may be able to circumvent the host restriction systems of *Propionibacterium*.

Vectors

A second aspect of the invention relates to a vector comprising a polynucleotide of the first aspect. The vector may be capable of replication in a host cell, such as a bacterium, for examples *Actinomycetes*, e.g. *Propionibacterium* or *E. coli*. The vector may be a linear polynucleotide or, more usually, a circular polynucleotide. The vector may be a hybrid of the polynucleotide of the invention and another vector. The other vector may be an *E. coli* vector, such as pBR322, or a vector of the pUC family, R1, ColD or RSF1010 or a vector derived therefrom.

The polynucleotide or vector of the invention may be a plasmid. Such a plasmid may have a restriction map the same as or substantially similar to the restriction maps shown in Figure 1,2a or 2b.

The polynucleotide or vector may have a size of 1 kb to 20 kb, such as from 2 to 10 kb, optimally from 3 to 7 kb.

The polynucleotide or vector may comprise multiple functional cloning sites. Such cloning sites generally comprise the recognition sequences of restriction enzymes. The polynucleotide or vector may comprise the sequence shown in SEQ ID No: 1 and/or contains restriction enzyme recognition sites for *EcoRI*, *SacI*, *AlwN1*, *BsmI*, *BsaBI*, *BclI*, *ApaI*, *HindIII*, *Sall*, *HpaI*, *PstI*, *SphI*, *BamHI*, *Acc65I*, *EcoRV* and *BglII*. The polynucleotide or vector may thus comprise one, more than one or all of these restriction enzyme sites, generally in the order shown in the Figures.

Preferably, when present in a bacterium, such as a *Propionibacterium* or *E. coli*, the polynucleotide or vector of the invention does not integrate into the chromosome of the bacterium. Generally the polynucleotide or vector does not integrate within 5 generations, preferably 20 or 30 generations.

The polynucleotide or vector may be an autonomously replicating plasmid that can remain extrachromosomal inside a host cell, which plasmid is derived from an endogenous *Propionibacterium* plasmid, and when comprising a heterologous gene (to the host) is capable of expressing that gene inside the host cell. The term "derived

from" means that the autonomously replicating plasmid includes a sequence the same as the polynucleotide of the invention.

The vector of the invention may comprise a selectable marker. The selectable marker may be one which confers antibiotic resistance, such as ampicillin, kanamycin or tetracylin resistance genes. The selectable marker may be an erythromycin resistance gene. The erythromycin resistance gene may be from *Actinomycetes*, such as *Saccharopolyspora erythraea*, for example from *Saccharopolyspora erythraea* NRRL2338. Other selectable markers which may be present in the vector include genes conferring resistance to chloramphenicol, thiostrepton, viomycin, neomycin, apramycin, hygromycin, bleomycin or streptomycin.

The vector may be an expression vector, and so may comprise a heterologous gene (which does not naturally occur in the host cell, e.g. *Propionibacteria*), or an endogenous or homologous gene of the host cell, e.g. *Propionibacteria*. In the expression vector the gene to be expressed is usually operably linked to a control sequence which is capable of providing for the expression of the gene in a host cell.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A controlled sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The heterologous or endogenous gene may be inserted between nucleotides 1 and 200 or between nucleotides 1500 to 3555 of SEQ ID No: 1 or at an equivalent position in a homologous polynucleotide.

Such genes may comprise homologous or endogenous genes such as for elongation factors, promoters regulatory sequences or elements, and replication proteins. Other genes (which may be heterologous to the host) include those encoding for or assisting in the production of nutritional factors, immunomodulators, hormones, proteins and enzymes (e.g. proteases, amylases, peptidases, lipases), texturing agents, flavouring substances (e.g. diacetyl, acetone), gene clusters, antimicrobial agents (e.g. nisin), substances for use in foodstuffs (e.g. in sausages, cheese) metabolic enzymes, vitamins (e.g. B₁₂), uroporphyrinogen (III)

methyltransferase (UP III MT), *cobA*, antigens and (e.g. for vaccines) therapeutic agents. As will be seen, the hosts can produce a wide variety of substances, not just polypeptides, which may be either the desired product or may be used to produce the desired product.

5 The heterologous gene may have a therapeutic effect on a human or animal. Such a gene may comprise an antigen, for example from a pathogenic organism. The host, such as *Propionibacterium*, comprising a polynucleotide with such a heterologous gene may be used as or in a vaccine, and may provide protection against the pathogens.

10 The heterologous antigen may be a complete protein or a part of a protein containing an epitope. The antigen may be from a bacterium, a virus, a yeast or a fungus.

Host cells and expression

15 The host cell forms the third aspect of the invention and comprises a polynucleotide or vector of the first or second aspect. The host cell may be a bacterium e.g. of the class *Actinomycetes*. The bacterium may be a *Propionibacterium* or *E. coli*. The *Propionibacterium* may be *P. freudenreichii*, *P. jensenii*, *P. thoenii* or *P. acidipropionici*.

20 In a fourth aspect the invention provides a process for producing a host cell of the third aspect, the process comprising transforming or transfecting a host cell with a polynucleotide or vector of the first or second aspect, e.g. with known transformation techniques¹¹.

25 In a fifth aspect the invention provides a process for the preparation of a polypeptide encoded by the polynucleotide or vector of invention present in host cell of the invention comprising placing or culturing the host cell in conditions where expression of the polypeptide occurs.

30 This aspect of the invention thus provides a process for the preparation of a polypeptide encoded by a given gene, which process comprises cultivating a host cell transformed or transfected with an expression vector comprising the gene, under conditions to provide for an expression of the said polypeptide, and optionally recovering the expressed polypeptide. The host cell may be of the class

Actinomycetes, or a gram positive bacteria such as *Propionibacterium* or *E. coli*.

Promoters, translation initiators, translation terminators, elongation factor genes, ribosomal RNA, antibiotic resistance genes, synthetic promoters (e.g. designed on consensus sequences) to other expression regulation signals present in the polynucleotide or vector can be those which are compatible with expression in the host cell. Such promoters include the promoters of the endogenous genes of the host cell.

Culturing conditions may be aerobic or anaerobic conditions, depending on the host. For a fermentation process the host cell would be placed in anaerobic, and then possibly aerobic, conditions. The compound produced, such as an expressed polypeptide, may then be recovered, e.g. from the host cell or fermentation medium. The expressed polypeptide may be secreted from the host cell. Alternatively the polypeptide may not be secreted from the host cell. In such a case the polypeptide may be expressed on the surface of the host cell. This may be desirable, for example, if the polypeptide comprises an antigen to which an immune response is desired in human or animal.

A homologous gene that may be present in the vector of the invention may be *cobA*. A host cell comprising this vector may therefore be capable of producing a compound such as vitamin B₁₂ from a substrate or the compound may be the product of an enzyme. The invention specifically provides a process for the preparation of vitamin B₁₂ comprising cultivating or fermenting such a host cell under conditions in which the UP(III) MT gene is expressed. The expressed enzyme can be contacted with a suitable substrate under conditions in which the substrate is converted to vitamin B₁₂. This may result in increased production of vitamin B₁₂.

Therapeutics

As described above the polynucleotide of the invention may comprise a heterologous gene which is a therapeutic gene. Thus the invention includes a host cell comprising a vector of the invention which comprises a therapeutic gene for use in a method of treatment of the human or animal body by therapy. Such a host cell may be *Propionibacterium*. The host cell may be alive or dead.

The host cell can be formulated for clinical administration by mixing them with

a pharmaceutically acceptable carrier or diluent. For example they can be formulated for topical, parenteral, intravenous, intramuscular, subcutaneous, oral or transdermal administration. The host cell may be mixed with any vehicle which is pharmaceutically acceptable and appropriate for the desired route of administration.

5 The pharmaceutically acceptable carrier or diluent for injection may be, for example, a sterile or isotonic solution such as Water for injection or physiological saline.

The dose of the host cells may be adjusted according to various parameters, especially according to the type of the host cells used, the age, weight and condition of the patient to be treated; the mode of administration used; the condition to be
10 treated; and the required clinical regimen. As a guide, the number of host cells administered, for example by oral administration, is from 10^7 to 10^{11} host cells per dose for a 70 kg adult human.

The routes of administration and dosages described are intended only as a guide since a skilled practitioner will be able to determine readily the optimum route of
15 administration and dosage of any particular patient and condition.

Polypeptides

A sixth aspect of the invention provides a polypeptide of the invention comprising one of the amino acid sequences set out in SEQ ID NO: 2 or 3 or a substantially homologous sequence, or of a fragment of either of these sequences.

20 The polypeptide may be one encoded by a polynucleotide of the first aspect. In general, the naturally occurring amino acid sequences shown in SEQ ID NO: 2 or 3 are preferred. However, the polypeptides of the invention include homologues of the natural sequences, and fragments of the natural sequences and their homologues, which have the activity of the naturally occurring polypeptides. One such activity
25 may be to effect the replication of the polynucleotide of the invention. In particular, a polypeptide of the invention may comprise:

- (a) the protein of SEQ ID No: 2 or 3; or
- (b) a homologue thereof from *Actinomycetes*, such as *Propionibacterium freudenreichii* or other *Propionibacterium* strains; or
- 30 (c) a protein at least 70% homologous to (a) or (b).

A homologue may occur naturally in a *Propionibacterium* and may function in a

substantially similar manner to a polypeptide of SEQ ID NO: 2 or 3. Such a homologue may occur in *Actinomyces* or gram positive bacteria.

A protein at least 70% homologous to the proteins of SEQ ID NO: 2 or 3 or a homologue thereof will be preferably at least 80 or 90% and more preferably at least 95%, 97% or 99% homologous thereto over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous amino acids. Methods of measuring protein homology are well known in the art and it will be understood by those of skill in the art that in the present context, homology is calculated on the basis of amino acid identity (sometimes referred to as "hard homology").

The sequences of the proteins of SEQ ID NO: 2 and 3 and of homologues can thus be modified to provide other polypeptides within the invention.

Amino acid substitutions may be made, for example from 1, 2 or 3 to 10, 20 or 30 substitutions. The modified polypeptide generally retains its natural activity. Conservative substitutions may be made, for example according to the following Table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

Polypeptides of the invention also include fragments of the above-mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID NO: 2 or 3. Such fragments can retain the natural activity of the full-length polypeptide.

Suitable fragments will be at least about 5, e.g. 10, 12, 15 or 20 amino acids in size. Polypeptide fragments of SEQ ID No: 2 and 3 and homologues thereof may contain one or more (e.g. 2, 3, 5, or 10) substitutions, deletions or insertions,

including conserved substitutions.

Polypeptides of the invention may be in a substantially isolated form. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than
5 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I , ^{35}S , enzymes, antibodies,
10 polynucleotides and linkers such as biotin.

Industrial Applications

As will be apparent from the discussion, the host cells of the third aspect can be used to produce not only the recombinant proteins, but also other compounds of interest, including non-proteins such as inorganic chemicals, in particular vitamins.
15 A seventh aspect of the present invention therefore relates to a process for the production of a compound, the process comprising culturing or fermenting host cells of the third aspect under conditions whereby the desired compound is produced. Although this compound may be a polypeptide, for example a polypeptide of the sixth aspect, it may also be one of the compounds mentioned in the previous
20 discussion concerning genes to be expressed. Clearly inorganic compounds will not be expressed by a gene, but they may be produced by an enzyme, or the polypeptide or enzyme may assist the host cell in the production of the desired compound. These compounds may be produced inside the cell, and later isolated, for example following lysis of the host cell, or they may pass through the wall of the host cell into
25 a surrounding medium, which may be a fermentation medium, for example an aqueous solution. In this way, the host cells can be cultured in an aqueous medium that comprises cells and nutrients for the cells, for example assimilable sources of carbon and/or nitrogen.

The polypeptides so produced may have therapeutic uses. They may be drugs or
30 other pharmacologically active compounds, or may be antigenic or immunogenic, in which case they may find use in vaccines.

The invention additionally encompasses compounds produced by this process, whether or not it is a recombinant polypeptide. Compounds specifically contemplated are vitamins, such as vitamin B₁₂ (cobalamin).

5 In some cases the compound need not be isolated either from the fermentation medium or from the host cells. The host cells may themselves be used in particular applications, for example in, or in the manufacturing of, foodstuffs such as sausages, or in cheese making, or the host cells may for example be included in an animal feed, such as when the host cells contain a compound to be ingested by the animal in question. The invention therefore extends to the use of these compounds or the host
10 cells, in the production of foodstuffs such as cheeses and sausages. The invention also contemplates foodstuffs or animal feed comprising host cells or a compound produced by the invention.

In a particularly preferred embodiment of the present invention the host cells can be used in a cheese making process, and so the invention additionally includes a
15 process for manufacturing cheese where the microorganisms employed are host cells of the invention. The host cells may be used instead of or in addition to, other bacteria, such as lactic acid bacteria. Propionic acid bacteria are currently used in cheese making processes, for example with mesophilic cultures (Maasdam type of cheese) as well as thermophilic cultures (Emmental). Both mesophilic and
20 thermophilic organisms can be responsible for the acidification of the milk or cheese. In this way the host cells of the invention can be not only used for cheese but also for the production of other fermented dairy products (e.g. yoghurts). Propionic acid bacterium are employed in cheese making because of their ability to convert lactate and carbohydrates to propionic acid, acetic acid and carbon dioxide. The host cells
25 of the invention, especially if they are propionibacteria, can be employed because they can be less sensitive to nitrates and salt, which may allow the reduction or omission of bacto-fugation of the milk (usually employed to reduce the levels of *Clostridia*).

The fermentation of the host cells may have one or two phases or stages. These
30 may be for example a growth and/or production phase, or anaerobic and/or aerobic phase. Preferably, there will be a growth and/or anaerobic phase, and suitably also (e.g. afterwards) a production and/or aerobic phase.

Both the carbon and/or nitrogen sources may be complex sources or individual compounds. For carbon, it is preferred that this is glucose. For nitrogen, appropriate sources include yeast extract or ammonia or ammonium ions.

Preferred features and characteristics of one aspect of the invention are suitable
5 for another aspect mutatis mutandis.

Figures

The invention is illustrated by the accompanying drawings in which:

Figure 1 is a restriction map of a vector within the invention, p545 obtained
from *P. freudenreichii* LMG 16545 (CBS 101022); and

10 Figures 2a and 2b each contain two maps of two vectors, all four vectors being
within the invention.

The invention will now be described, by way of example, by reference to the
following Examples, which are not to be construed as being limiting.

EXAMPLE 1

Screening of *Propionibacterium* strains

A collection of 75 nonpathogenic strains of *Propionibacterium* was screened for the presence of indigenous plasmids. The majority of strains were obtained from the BCCM/LMG culture collection (Ghent, Belgium), although some strains were obtained from ATCC (Rockville, Md., USA) or from DSM (Braunschweig, Germany). Screening was performed using a small scale plasmid isolation procedure. First bacteria were cultivated anaerobically in MRS medium⁶ for 48 hrs at 30°C.

Plasmids were then purified from the bacteria using a plasmid DNA isolation procedure originally developed for *E. coli*⁴ with some modifications: cells from a 5 ml culture were washed in a 25% sucrose, 50 mM Tris-HCl pH8 solution, resuspended in 250µl TENS (25% sucrose + 50mM NaCl + 50 mM Tris-HCl + 5mM EDTA pH8), containing 10mg/ml lysozyme (Boehringer Mannheim), and incubated at 37°C for 20-30 minutes. The bacterial cells were then lysed in 500µl of 0.2 N NaOH/1% SDS (2-5 minute incubation on ice). After addition of 400µl 3M NaAc pH4.8 (5 minutes on ice) and subsequent extraction with phenol/chloroform, the DNA was precipitated by addition of isopropanol.

The DNA was analysed by electrophoresis on 1% agarose gels, and visualised by ethidium bromide. Whereas most strains were negative, i.e. did not reveal the presence of indigenous plasmids in this analysis, the majority of strains that proved positive contained large (≥ 20 kb) plasmids. Smaller plasmids were observed in 6 strains. Of these, *P. jensenii* LMG16453, *P. acidipropionici* ATCC4875, *P. acidipropionici* LMG16447 and a nonspecified *Propionibacterium* strain (LMG16550) contained a plasmid in the size range of 6-10 kb. Two strains (*P. freudenreichii* LMG16545 and *P. freudenreichii* LMG16546) showed an identical plasmid profile of 2 plasmids. One plasmid was large (size not determined) and the other was smaller, more abundantly present and had a size of 3.6 kb. These 3.6 kb plasmids from LMG16545 and LMG16546 were chosen for further analysis.

EXAMPLE 2

Analysis of an indigenous plasmid from strains LMG16545 and LMG16546

The 3.6 kb plasmids were isolated from both strains and further purified by CsCl-ethidium bromide density gradient ultracentrifugation¹¹. Limited restriction maps were made of both preparations and these turned out to be identical¹¹. The restriction map of the 3.6 kb plasmid is shown in Fig.1. Restriction enzymes and T4 ligase were obtained from New England Biolabs or GIBCO BRL.

The 3.6 kb plasmid from strain LMG16545 (from here on referred to as p545) was radioactively labelled and used in Southern blot hybridization experiments.

Hybridisation conditions were 0.2 x SSC, 65°C. It reacted equally well with both LMG16545 and LMG16546 plasmid DNA extracts, supporting the close relationship of these strains, whereas a plasmid DNA extract from *P.acidipropionici* ATCC4875, that harbors a 6.6 kb plasmid called pTY1 or pRG01⁸ failed to react.

The DNA sequence of plasmid p545 was determined with fluorescent dye labelled dideoxyribonucleotides in an Applied Biosystems 373A automatic sequencer, and is included as SEQ ID No: 1 in the sequence listing. Sequence analysis was performed on plasmid DNA that had been linearized with *EcoRI* and inserted into *EcoRI* digested pBluescript SKII+ DNA (Stratagene, La Jolla, Ca., USA). Computer assisted analysis of the sequence thus obtained using BLAST search¹ revealed homologies to proteins involved in replication of plasmids from several GC-rich organisms (e.g., pAL5000 encoded *repA* and *repB* from *Mycobacterium fortuitum*^{8,14} show 28-30% identity and 34-38% similarity with the respective putative replication proteins from plasmid p545; pXZ10142 from *Corynebacterium glutamicum* [PIR Accession Number S32701] is another example of plasmids encoding replication proteins homologous to the p545 putative replication proteins). The results of the database comparisons with homologous sequences are detailed in Examples 7 and 8.

EXAMPLE 3

Construction of *E. coli*/Propionibacterium shuttle vectors

E. coli plasmid pBR322 was digested with *EcoRI* and *AvaI* and the smaller

fragment thus generated (measuring 1.4 kb and encompassing the tetracyclin resistance conferring gene) was replaced by a synthetic duplex DNA. The synthetic duplex DNA was designed so as to link *EcoRI* and *AvaI* ends and to supply a number of unique restriction enzyme recognition sites:

5 5' -
AATTCAAGCTTGTCTGACGTTAACCTGCAGGCATGCGGATCCGGTACCGAT
ATCAGATCT - 3' (SEQ. ID. NO. 4)
3' -
GTTCTGAACAGCTGCAATTGGACGTCCGTACGCCTAGGCCATGGCTATAGT
10 CTAGAAGCC - 5' (SEQ.ID.NO. 5)

The following restriction enzyme recognition sites were supplied in this way:
EcoRI (restored), *HindIII*, *SalI*, *HpaI*, *PstI*, *SphI*, *BamHI*, *Acc65I*, *EcoRV*, *BglII*
(*AvaI* is not restored).

15 This synthetic DNA was ligated to the large fragment and the ligation mixture transferred back to *E. coli* (T4 ligase was used). A plasmid of the expected composition was obtained (pBR322ΔI). The multiple cloning site can be used to introduce a selection marker as well as plasmid p545 DNA.

As an example the construction of an *E. coli*/*Propionibacterium* shuttle plasmid conferring resistance to erythromycin was performed as will now be described.
20 A 1.7 kb *Acc65I* fragment from the *Saccharopolyspora erythraea* NRRL2338 erythromycin biosynthesis cluster and containing the erythromycin resistance conferring gene^{15,2} was inserted into *Acc65I* linearized pBR322ΔI. Then the newly derived construct, named pBRES, was linearized with *EcoRV* and ligated to p545 DNA that had been digested with *BsaBI*. *E. coli* transformants were found to harbor a
25 vector with the correct insert, in both orientations. The resulting plasmid vectors were named pBRESP36B1 and pBRESP36B2 (Figs. 2a and 2b).

Plasmid vector constructs were also obtained with p545 DNA linearized in an other restriction site situated outside the putative replication region, namely *AlwNI*. For this construction the pBRES vector had to be provided with a suitable cloning
30 site. An adaptor was designed consisting of two complementary oligonucleotides of

the following composition (SEQ. ID. Nos 6 and 7):

5' GTACCGGCCGCTGCGGCCAAGCTT 3'

5' GATCAAGCTTGGCCGCAGCGGCCG 3'

Annealing of these oligo's created a double stranded DNA fragment with *Acc65I* and *BglII* cohesive ends respectively, which moreover contains an internal *SfiI* restriction site, that provides ends compatible to the *AlwNI* digested p545 plasmid. This adaptor was cloned in pBRES between the *BglII* and the proximal *Acc65I* site. The pBRES-Sfi vector thus obtained was subsequently digested by *SfiI* and ligated to *AlwNI* digested p545. Transformation of *E.coli* yielded transformants with the correct vector as confirmed by restriction enzyme analysis. The vector obtained was named pBRESP36A (Fig.2).

EXAMPLE 4

Transformation of *Propionibacterium* with *E. coli*/ *Propionibacterium* shuttle vectors

Transformation of *Propionibacterium freudenreichii* strain ATCC6207 with pBRESP36B1 will be described.

The bacterial cells are cultivated in SLB (sodium lactate broth¹⁷ at 30°C to a stationary growth phase, and subsequently diluted 1:50 in fresh SLB. After incubation at 30°C for around 20 hours, cells (now in the exponential growth phase) were harvested and washed extensively in cold 0.5M sucrose. Subsequently cells were washed once in the electroporation buffer, consisting of 0.5M sucrose, buffered by 1mM potassiumacetate, pH5.5, and finally resuspended in this electroporation buffer in about 1/100 of the original culture volume. Cells were kept on ice during the whole procedure.

For the electroporation (apparatus from BIORAD), 80 - 100 µl of cell suspension was mixed with ±1 µg of DNA (or smaller amounts), in a cooled 1 or 2 mm electroporation cuvette, and an electric pulse delivered. Optimal pulse conditions were found to be 25kV/cm at 200 Ω resistance and 25µF capacitance. However, lower and higher voltages (also at 100Ω) also yield transformants.

Immediately after the pulse, 900 µl cold SLB containing 0.5 M sucrose was added to the pulsed cell suspension and these are subsequently incubated for 2.5 to 3 hours at 30°C before plating appropriate dilutions on SLB/agar plates containing 0.5 M sucrose and 10µg/ml erythromycin. After a 5 to 7 day incubation period at 30°C under anaerobic conditions, transformants were detected.

DNA isolated from *E. coli* DH5α (Promega) yielded a transformation efficiency of 20 - 30 transformants per µg DNA. A 10-100 fold higher efficiency is achieved when DNA is isolated from *E. coli* JM110 (dam⁻, dcm⁻ strain). *E. coli* transformation was done according to BIORAD instructions.

Transformants contained the authentic vectors, indistinguishable from the original plasmid DNA used for transformation of ATCC6207. This was shown by restriction enzyme analysis of plasmid DNA isolated from the transformants by the small scale plasmid DNA isolation procedure referred to before.

Vectors were exclusively present as autonomously replicating plasmids.

Southern blot hybridization¹³ with total DNA isolates showed that chromosomal DNA did not hybridise to the vector DNA used as a probe, indicating that no chromosomal integration of plasmid DNA occurred.

Transformation was also successful with vectors pBRESP36B2 and pBRESP36A, indicating that functionality of the vector was independent of the orientation of p545 or the cloning site used. Also in this case the authenticity of the vectors was confirmed.

Moreover, transformation of *P. freudenreichii* strain ATCC6207 with DNA isolated from a *Propionibacterium* transformant resulted in a 10⁵-10⁶ fold increased transformation efficiency as compared to that obtained with DNA isolated from *E. coli* DH5α.

Transformation of another *P. freudenreichii* strain, LMG16545 (the same strain from which the p545 plasmid was obtained), resulted in a transformation efficiency comparable to that of the ATCC strain.

The results of the transformations, and the effect on vitamin B₁₂ production, is shown in the following Table.

Eight out of 10 transformants gave up to a 50% higher vitamin B₁₂ content than the control strain.

5	Strain ID	Transformed plasmid	Vitamin B ₁₂ (mg/g dry matter)
	COB1	pBRES36COB	0.57
	COB2	pBRES36COB	0.67
	COB3	pBRES36COB	0.83
	COB4	pBRES36COB	0.68
10	COB5	pBRES36COB	0.69
	COB6	pBRES36COB	0.61
	COB7	pBRES36COB	0.53
	COB8	pBRES36COB	0.64
	COB9	pBRES36COB	0.50
15	COB10	pBRES36COB	0.74
	recATCC6207	pBRESP36B2	0.54

EXAMPLE 5

Construction of plasmid vector containing the *cobA* gene

The construction and application of a plasmid vector to increase the level of vitamin B₁₂ (cobalamin) synthesis in *P. freudenreichii* strain ATCC6207 will be described.

The promoter region of the gene conferring erythromycin resistance in *Saccharopolyspora erythraea*^{2,3} was generated by PCR using the following primers (SEQ. ID. NOs 8 and 9):

forward primer: (5' - 3')

AAACTGCAGCTGCTGGCTTGCGCCCGATGCTAGTC

reverse primer: (5' - 3')

AAACTGCAGCAGCTGGGCAGGCCGCTGGACGGCCTGCCCTCGAGCTCGT
CTAGAATGTGCTGCCGATCCTGGTTGC

The PCR fragment thus generated contained an *Alw*NI site at the 5' end followed
5 by the authentic promoter region and the first 19 amino acids of the coding region of
the erythromycin resistance gene, to ensure proper transcription and translation
initiation. At the 3' end *Xba*I and *Xho*I sites were provided (for insertion of the *cobA*
gene in a later stage), a terminator sequence as present downstream from the
erythromycin resistance gene, and an *Alw*NI site.

10 The PCR product was digested by *Alw*NI and ligated to pBRESP36B2, partially
digested with *Alw*NI. Of the two *Alw*NI sites present in pBRESP36B2, only the one
present in the p545 specific part of the vector will accommodate the fragment. *E.*
coli transformants were obtained harboring the expected construct, named
pBRES36pEt. This vector was used for further constructions as described below.

15 The coding sequence of *cobA*, the gene encoding uroporphyrinogen III
methyltransferase, was generated by PCR from *Propionibacterium freudenreichii*
strain ATCC6207, using the following primers (SEQ. ID. NOs 10 and 11):

forward: (5'- 3')

CTAGTCTAGACACCGATGAGGAAACCCGATGA

20 reverse: (5'- 3')

CCCAAGCTTCTCGAGTCAGTGGTCGCTGGGCGCGCG

The *cobA* gene thus amplified carries an *Xba*I site at the N terminal coding
region, and *Hind*III and *Xho*I sites at the C terminal coding region.

The functionality of this *cobA* gene was confirmed by cloning the PCR product
25 as an *Xba*I -*Hind*III fragment in pUC18, and subsequent transformation of *E. coli*
strain JM109. Transformants with a functional *cobA* gene show a bright red
fluorescence when illuminated with UV light. Plasmid DNA isolated from such a
transformant was digested with *Xba*I and *Xho*I, ligated to likewise digested
pBRESP36B₂. DNA and used for transformation of *E. coli*. DNA from several
30 transformants was analysed by restriction enzyme digestion and gel electrophoresis.
Transformants were found to bear the correct insert in the expression vector. This

new vector was named pBRES36COB. This vector was subsequently transferred to *P. freudenreichii* ATCC6207 following the protocol described before. Ten of the transformants obtained were analysed and were found to harbor the pBRES36COB vector, which was again indistinguishable from the original vector used for transformation, as shown by analysis of the restriction enzyme digests. In these ten transformants the level of vitamin B₁₂ synthesis was determined as follows.

Frozen cultures of *Propionibacterium* transformants 1 through 10, as well as a control strain containing only the vector plasmid pBRESP36B₂, were inoculated in 100 ml flasks containing 50 ml of BHI (Brain Heart Infusion) medium (Difco) and incubated for 72 hrs at 28°C without shaking. From this preculture 4 ml were transferred to 200 ml of production medium consisting of Difco yeast extract 15 g/l, Nalactate 30 g/l, KH₂PO₄ 0.5g/l, MnSO₄ 0.01 g/l, and CoCl₂ 0.005 g/l in a 500 ml shake flask and incubated at 28°C for 56 hrs without shaking, followed by 48 hrs in a New Brunswick rotary shaker at 200 rpm.

Vitamin B₁₂ titres were measured using a known HPLC method⁵. Nine out of 10 transformants showed an approx. 25% higher vitamin B¹² production than the control strain.

EXAMPLE 6

Stability of the plasmids

All three shuttle vectors pBRESP36A, pBRESP36B1, and pBRESP36B2 were stably maintained over 30 generations of culturing of the respective transformants: no loss of erythromycin resistance was observed as determined by viability counts on selective (erythromycin containing) and non-selective agar plates. The structural stability of the plasmid in the transformant population after 30 generations was established by plasmid DNA isolation and characterisation by restriction enzyme mapping as described above: only restriction fragments similar to those of the authentic plasmid were observed.

EXAMPLE 7

Database sequence homology analysis for predicted polypeptide encoded by the first open reading frame (SEQ.ID. No. 2)

5 MDSFETLFPESWLPRKPLASAEKSGAYRHVTRQRALELPYIEANPLVMQSLV
ITDRDASDADWAADLAGLPSYVSMNRVTTTGHIVYALKNPVCLTDAARR
RPINLLARVEQGLCDVLGGDASYGHRITKNPLSTAHATLWGPADALYELRA
LAHTLDEIHALPEAGNPRRVNTRSTVGRNVTLFDTTRMWAYRAVRHSWGG
PVAEWEHTVFEHIHLLNETIID

10 The above 227 amino acid sequence (ORF1) was aligned and compared with several other protein sequences (target NBRF-PIR, release PIR R52.0 March 1997, cut-off 45, KTUP:2).

15 With a protein from *Mycobacterium fortuitum* plasmid pAL 5,000 (JS0052) a match of 37.1% was found over 194 amino acids (INIT 167,292). With a protein from *Corynebacterium glutamicum* (S32701) a match 32.0% over 125 amino acids was found (INIT 125, 116). A match of 29.9% over 221 amino acids (INIT 86, 259) was found with the Cole2 protein from *E. coli* (SO4455). Precisely the same match over the same number of amino acids was found for the Cole3 protein, also from *E. coli* (SO4456).

EXAMPLE 8

20 Database sequence homology analysis for predicted polypeptide encoded by the second reading frame (SEQ.ID. No. 3)

MTTRERLPRN GYSIAAAAKK LGVSESTVKR WTSEPREEFV ARVAARHARI
REL RSEGQSM RAIAAEVGVS VGT VH YALNK NRTDA

25 The second protein (ORF2) was also aligned and compared with another protein, using the same parameters and software as described for Example 7. This sequence however is only 85 amino acids in length.

The ORF2 sequence was compared with a protein from *Mycobacterium fortuitum* (S32702) and a 53.3% match over 75 amino acids was found (INIT 207, 207).

EXAMPLE 9

5 Functional analysis of plasmid p545

In order to improve the vector system further, the plasmid functions essential for replication and stability were delineated more precisely through deletion of large regions of the original plasmid p545. The result, a smaller cloning vector, will allow the use of the p545 based vector system for cloning larger DNA fragments.

10 To this end vector pBRESP36A (Figure 2) was digested with *Sst*II and *Bcl*II, resulting in a 1.7 kb and a 6.5 kb fragment. The 1.7 kb fragment, in fact the 1.6 kb *Alw*NI - *Bcl*II fragment of plasmid p545, was replaced by a synthetic duplex DNA, composed of SEQ. ID. No. 12 and SEQ. ID. No. 13 with *Sst*II and *Bcl*II compatible ends and a number of unique restriction enzyme recognition sites.

15 SEQ. ID. No. 12 5' GGAGATCTAGATCGATATCTCGAG 3'
SEQ. ID. No. 13 5' GATCCTCGAGATATCGATCTAGATCTCCGC 3'

The following restriction enzyme recognition sites were supplied in this way:

20 *Sst*II (restored), *Bgl*II, *Xba*I, *Cla*I, *Eco*RV, *Xho*I, (*Bcl*II is not restored). The ligation mixture was transferred to *E. coli*, and transformants were selected containing a vector of the expected composition. The vector was named pBRESAΔS-B. Subsequent successful transformation of *P. freudenreichii* strain ATCC6207 with this vector indicated that the 1.6 kb region between *Alw*NI and *Bcl*II in p545 is not essential for replication of the plasmid.

25 A further deletion was made by removal of the 240 bp corresponding to the region between *Sal*II and *Bcl*II in plasmid p545. This was achieved by digestion of pBRESAΔS-B with *Sal*II - *Sst*I, and *Sst*I - *Xho*I respectively, and isolation of the 1.3 kb *Sal*II - *Sst*I fragment, and the 6.6 kb *Sst*I - *Xho*I fragment. The fragments were ligated, and the ligation mixture was transferred to *P. freudenreichii* ATCC6207,
30 yielding numerous transformants. The newly derived construct, named

pBRESAΔS-S, was isolated and its structure confirmed by restriction enzyme mapping.

Thus all essential information for replication of plasmid p545 is located on a fragment of 1.7 kb delimited by the restriction sites *SalI* and *AlwNI* and encompassing the predicted replication proteins encoded by ORF1 and ORF2, and that 1.8 kb can be deleted without obviously disturbing replication or stability of the plasmid.

EXAMPLE 10

Expression of a chloramphenicol resistance gene from *Corynebacterium*.

A chloramphenicol resistance gene (*cml*) from *Corynebacterium*¹⁸ has been identified as encoding a chloramphenicol export protein. This gene was inserted in the *Propionibacterium* - *E. coli* shuttle vector pBRESP36B2. This vector was digested with *BglII* and *HindIII*, and with *BglII* and *HpaI* respectively. The 2.9 kb *BglII* - *HindIII* and 5.2 kb *BglII* - *HpaI* fragments were isolated.

The fragment containing the *cml* gene, including its own promoter, was obtained by digestion with *PvuII* and *HindIII*, and the 3.3 kb fragment containing the gene was isolated. The two vector-specific fragments and the *cml* fragment were ligated: *PvuII* and *HpaI* ends are blunt, thus inserting the *cml* gene as well as restoring the *ermE* gene of the parent vector. The ligation mixture was transferred to *E. coli*, and a transformant was selected, in which the vector contained the correct *cml* insert. The vector was named pBRESBCM.

Transformation of *P. freudenreichii* ATCC6207 with this vector, and selection on plates containing 10μg/ml erythromycin, or 5μg/ml chloramphenicol, yielded erythromycin and chloramphenicol resistant colonies, respectively, indicating that apart from the erythromycin resistance gene (shown earlier with the *Propionibacterium* - *E. coli* shuttle vectors), also the chloramphenicol resistance gene is expressed in *Propionibacterium*. Transformants could be cultivated in liquid medium containing up to 20 μg/ml chloramphenicol.

EXAMPLE 11

Expression of lipase (*gehA*) gene from *P. acnes*

To illustrate efficient cloning and expression of an extracellular protein using the present vector system, a lipase gene *gehA* from *P. acnes* was used¹⁹. Vector pUL6001, harbouring *gehA* on a *XhoI* fragment, was digested with *XhoI*, and the
5 2.75 kb fragment containing the gene was isolated. Vector pBRESAΔS-B (from Example 9) was linearized by *XhoI*, and the ends dephosphorylated using Calf Intestine Phosphatase to avoid self ligation. Linearized vector and the *gehA* containing fragment were ligated and the ligation mixture was transferred to *E. coli*.
10 Transformants were analysed by restriction enzyme analysis for the presence of the correct recombinant plasmid, named pBRESALIP. This plasmid was subsequently transferred to *P. freudenreichii* strain ATCC6207. Transformants were screened for the expression of the lipase gene, using agar plates containing tributyrin as the indicator for increased lipase expression. *P. freudenreichii* transformants harbouring
15 pBRESALIP showed significantly increased halo sizes in this assay as compared to untransformed strains or strains transformed with the parent vector.

REFERENCES

1. Altschul *et al*, J. Mol. Biol. 215: 403-410 (1990)
2. Bibb *et al.*, Gene 38, 215 (1985)
- 5 3. Bibb *et al.*, Mol. Microbiol. 14(3), 533 (1984)
4. Birnboim and Doly, Nucleic Acids Res. 7, 1513 (1979)
5. Blanche, Anal. Biochem. 189, 24 (1990)
6. DeMan *et al*, J. Appl. Bacteriol. 36, 130 (1960)
7. Deveraux *et al*, Nucleic Acids Research. 12:387-395 (1984)
- 10 8. Labidi *et al*, Plasmid 27, 130 (1992)
9. Rehberger and Glatz, Appl. Environ. Microbiol. 59, 83 (1990)
10. Rossi *et al*, Res. Microbiol. 147, 133 (1996)
11. Sambrook *et al*, Molecular Cloning, Cold Spring Harbor Laboratory Press (1989)
- 15 12. Sattler *et al*, J. Bact. 177, 1564 (1995)
13. Southern, J. Mol. Biol. 98, 503 (1975)
14. Stolt and Stoker, Microbiol. 142, 2795 (1996)
15. Thompson *et al*, Gene 20, 51 (1982)
16. Uchijama and Weisblum, Gene 38, 103 (1985)
- 20 17. de Vries *et al*, J. Gen. Microbiol. 71, 515 (1972)
18. Tauch *et al*, Plasmid 40(2): 126 (1998)
19. Miskin *et al*, Microbiol 143: 1745 (1997)

SUMMARY OF SEQUENCES

1. DNA sequence of plasmid LMG 16545 (CBS 101022), 3.6 kb.
- 25 2. amino acid of protein of ORF1 (303 residues, bases 273-1184).
3. amino acid of protein of ORF2 (85 residues, bases 1181-1438).
- 4-13. DNA primers/oligonucleotides

CLAIMS

1. A polynucleotide comprising a sequence capable of hybridising selectively to
 - (a) SEQ ID NO: 1 or the complement thereof;
 - 5 (b) a sequence from the 3.6 kb plasmid of *Propionibacterium freudenreichii* CBS 101022;
 - (c) a sequence from the 3.6 kb plasmid of *Propionibacterium freudenreichii* CBS 101023; or
 - (d) a sequence that encodes a polypeptide which comprises a SEQ. ID. No. 2
10 or 3, an amino acid sequence substantially homologous thereto or a fragment of either sequence.
2. A polynucleotide which is an autonomously replicating plasmid that can remain extrachromosomal inside a host cell, which plasmid is derived from an endogenous *Propionibacterium* plasmid, and when comprising a heterologous gene
15 is capable of expressing that gene inside the host cell.
3. A polynucleotide according to claim 1 which is autonomously replicating in a host cell.
4. A polynucleotide according to claim 3 in which the host cell is a *Propionibacterium*.
- 20 5. A polynucleotide according to claim 4 in which the *Propionibacterium* is *Propionibacterium freudenreichii*.
6. A polynucleotide according to any one of the preceding claims which is capable of selectively hybridising to one or more sequence(s) in SEQ ID No:1 which is (or are) necessary for autonomous replication in a *Propionibacterium*.
- 25 7. A polynucleotide according to claim 1 which comprises either the 1.7 kb fragment of SEQ. ID. No. 1 delineated by restriction sites *SalI* and *AlwNI* or nucleotides 1 to 1750 of SEQ. ID. No. 1.
8. A vector which comprises a polynucleotide according to any one of the preceding claims.
- 30 9. A vector according to claim 8 which is a plasmid.
10. A vector according to claim 8 or 9 which additionally comprises a

selectable marker.

11. A vector according to any one of claims 8 to 10 which is autonomously replicating in *E. coli*.

5 12. A vector according to any one of claims 8 to 11 which is an expression vector.

13. A vector according to claim 12 which comprises an endogenous gene of a *Propionibacterium* or a heterologous gene operatively linked to a control sequence which is capable of providing for expression of the gene.

14. A vector according to claim 13 in which the gene is the *cobA* gene.

10 15. A vector according to claim 13 in which the heterologous gene encodes a polypeptide which is therapeutic in a human or animal.

16. A polypeptide which comprises the sequence SEQ ID No: 2 or 3 or a sequence substantially homologous thereto, or a fragment of either said sequence, or is encoded by a polynucleotide as defined in any of claims 1 to 7.

15 17. A host cell comprising a heterogeneous polynucleotide or vector according to any one of claims 1 to 15 or which can be transformed or transfected with a vector according to any one of claims 13 to 15.

18. A host cell according to claim 17 which is a bacterium.

20 19. A host cell according to claim 18 which is a *Propionibacterium* or *E. coli*.

20. A process for producing a host cell according to any one of claims 17 to 19 comprising transforming or transfecting a host cell with a polynucleotide or vector according to any one of claims 1 to 15.

25 21. A process for the preparation of a polypeptide, or other compound, the process comprising cultivating or fermenting a host cell as defined in any one of claims 17 to 19 under conditions that allow expression or production of the polypeptide or compound.

22. A process according to claim 21 which is a fermentation process wherein the host cell is cultured in aerobic or anaerobic conditions.

30 23. A process according to claim 21 or 22 in which the expressed polypeptide or produced compound is recovered from the host cell.

24. A process according to claim 23 wherein the polypeptide is a protease,

amylase, lipase or peptidase or the compound is vitamin B₁₂.

25. A process according to any one of claims 21 to 24 where the polypeptide is secreted from the host cell.

5 26. A process according to claim 25 in which the polypeptide is expressed on the surface of the host cell and/or the polypeptide is an antigen or immunogen.

27. A polypeptide or compound prepared by a process according to any one of claims 20 to 26.

10 28. A process for the production of vitamin B₁₂ (cobalamin), the process comprising culturing a host cell according to any one of claims 17 to 19 under conditions in which the vitamin is produced and, if necessary, isolating the vitamin.

29. Vitamin B₁₂ produced by a process according to claim 28.

30. A polypeptide according to claim 27 for use in a method of treating the human or animal body by therapy.

15 31. A host cell according to any one of claims 17 to 19 for use in a method of treating the human or animal body by therapy or for use in an animal feed.

32. Use of a host cell according to any one of claims 17 to 19 or a polypeptide or compound according to claim 27 to either make cheese or for use in cheesemaking.

20 33. Use of a host cell according to any one of the claims 17 to 19 or a polypeptide or compound according to claim 27, in the manufacture of a foodstuff or in an animal feed.

34. A foodstuff comprising a polypeptide or compound according to claim 27 or a host cell according to any of claims 17 to 19.

25 35. A foodstuff according to claim 34 for consumption by humans (e.g. a cheese, sausage) or by an animal.

36. A process for manufacturing cheese or other fermented dairy product the process comprising using a host cell according to any of claims 17 to 19.

30 37. A process according to claim 36 wherein the host cell is used instead of or in addition to lactic acid bacteria.

38. A process according to claim 36 or 37 wherein the host cell is a *Propionibacterium* cell.

Figure 1

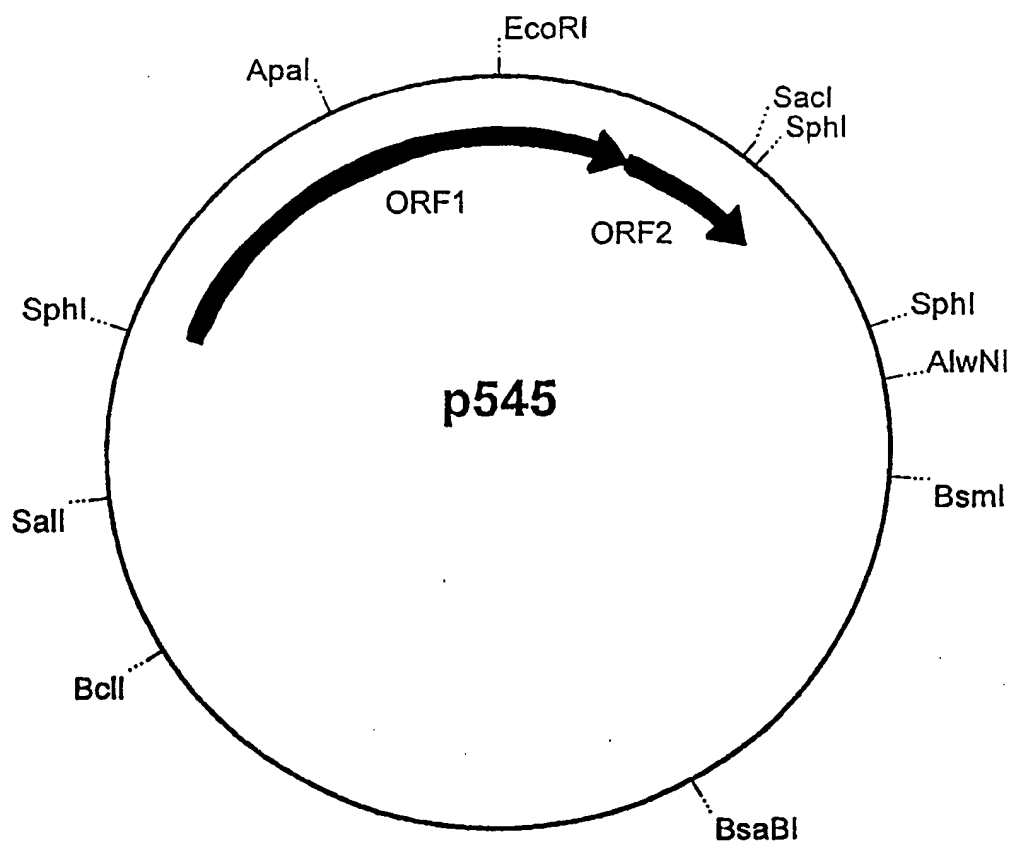


Figure 2a

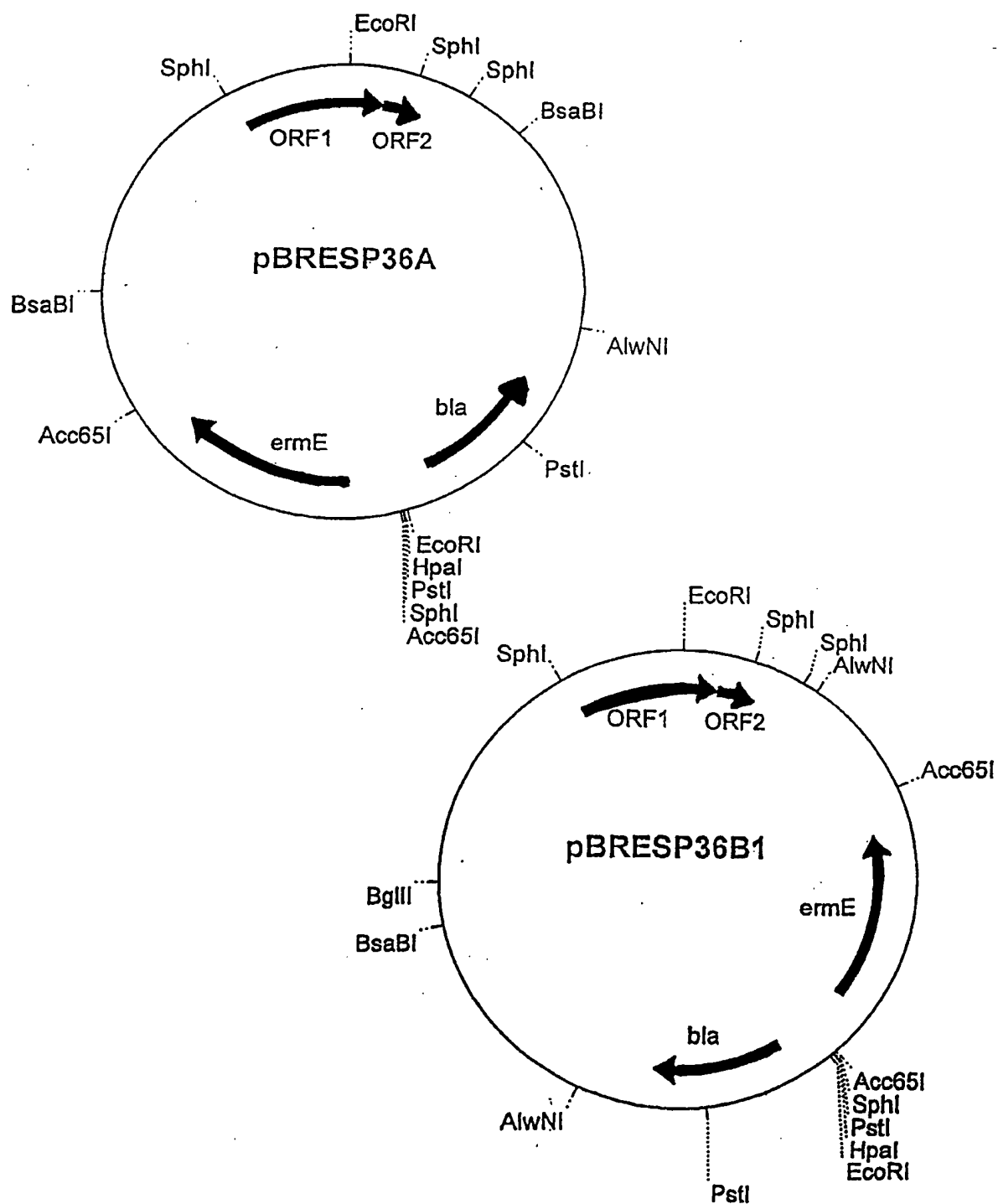
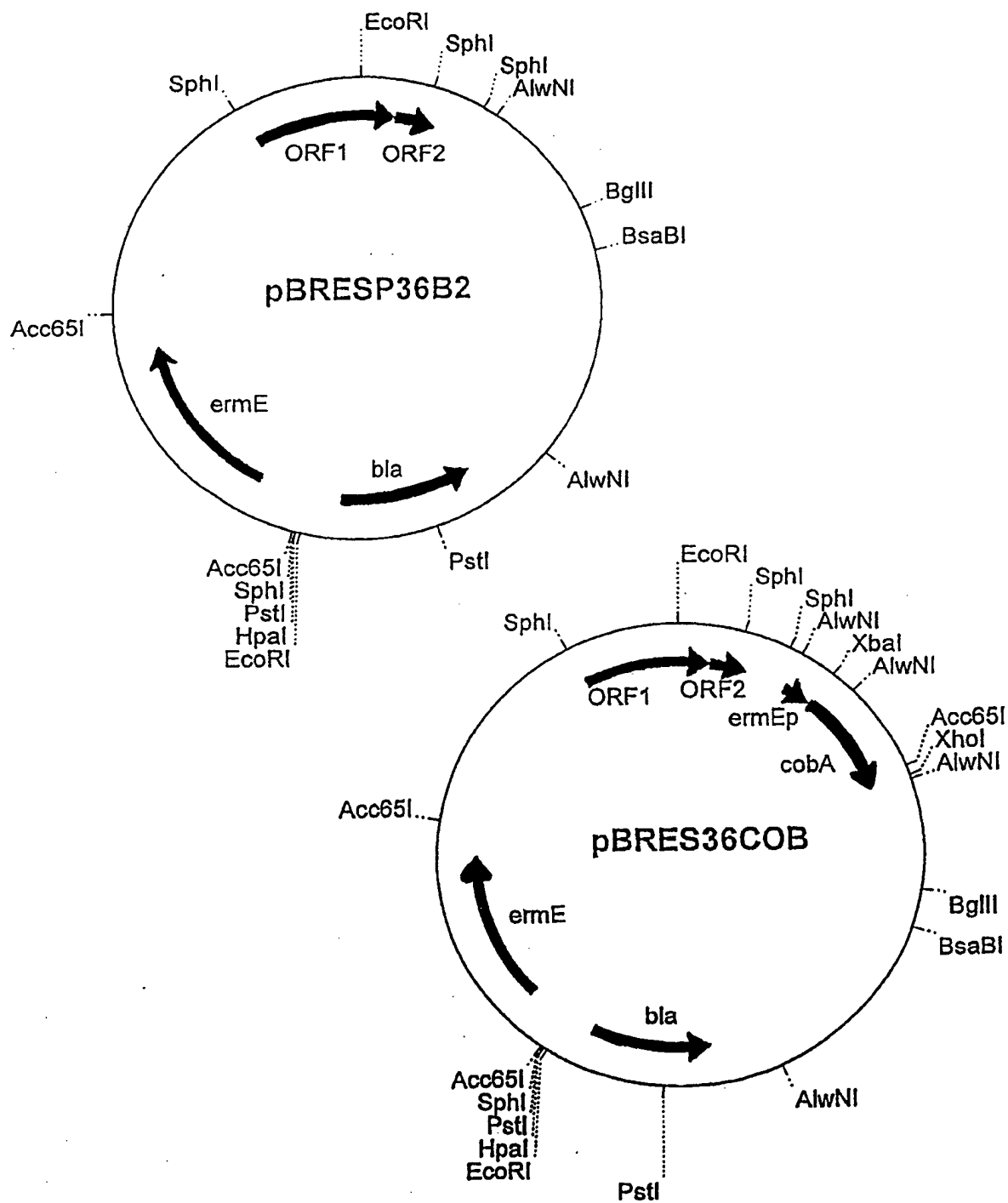


Figure 2b



SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Gist-brocades B.V.
(B) STREET: Wateringseweg 1
(C) CITY: Delft
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): 2611 XT

(ii) TITLE OF INVENTION: Propionibacterium Vector

(iii) NUMBER OF SEQUENCES: 13

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3555 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Propionibacterium freudenreichii
(C) INDIVIDUAL ISOLATE: CBS101022 LMG16545

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 273..1184
(D) OTHER INFORMATION: /gene= "ORF1"

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 1181..1438
(D) OTHER INFORMATION: /gene= "ORF2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GTCGACCCTG ACAGCCGGCG AGCAGTTCAG GCGAAGATCG CACAGCTGCG CGAGGAACTA

GCCGCAATGC CCGAACACGC CCCAGCCATC CCTTGGAGCA GGTGGCAGCG TCAGGGGAGT
 120
 CGGGGGATGT TTGGCAGGGG ATGTGGAAAG AGAGTTCGCT TTGCTCACAT GGCTCAACCG
 180
 5 GGTAACCTAAC TGATATGGGG TCTTCGTCGC CCACTTTGAA CACGCCGAGG AATGGACCAC
 240
 GCTGAACGTG ACTCGCATGC TTCACTGCAT GT ATG GAT TCG TTC GAG ACG TTG
 293
 10 Met Asp Ser Phe Glu Thr Leu
 1 5
 TTC CCT GAG AGC TGG CTG CCA CGC AAG CCG CTG GCG TCA GCC GAG AAG
 341
 Phe Pro Glu Ser Trp Leu Pro Arg Lys Pro Leu Ala Ser Ala Glu Lys
 10 15 20
 15 TCT GGG GCG TAC CGG CAC GTG ACT CGG CAG AGG GCG CTG GAG CTG CCT
 389
 Ser Gly Ala Tyr Arg His Val Thr Arg Gln Arg Ala Leu Glu Leu Pro
 25 30 35
 TAC ATC GAA GCG AAC CCG TTG GTC ATG CAG TCC TTG GTC ATC ACC GAT
 437
 20 Tyr Ile Glu Ala Asn Pro Leu Val Met Gln Ser Leu Val Ile Thr Asp
 40 45 50 55
 CGA GAT GCT TCG GAT GCT GAC TGG GCC GCA GAC CTC GCT GGG CTG CCT
 485
 25 Arg Asp Ala Ser Asp Ala Asp Trp Ala Ala Asp Leu Ala Gly Leu Pro
 60 65 70
 TCA CCG TCC TAC GTG TCC ATG AAC CGT GTC ACG ACC ACC GGA CAC ATC
 533
 30 Ser Pro Ser Tyr Val Ser Met Asn Arg Val Thr Thr Thr Gly His Ile
 75 80 85
 GTC TAT GCC TTG AAG AAC CCT GTG TGT CTG ACC GAT GCC GCG CGG CGA
 581
 Val Tyr Ala Leu Lys Asn Pro Val Cys Leu Thr Asp Ala Ala Arg Arg
 90 95 100
 35 CGG CCT ATC AAC CTG CTC GCC CGC GTC GAG CAG GGC CTA TGC GAC GTT
 629
 Arg Pro Ile Asn Leu Leu Ala Arg Val Glu Gln Gly Leu Cys Asp Val
 105 110 115
 40 CTC GGC GGC GAT GCA TCC TAC GGG CAC CGG ATC ACA AAG AAC CCG CTC
 677
 Leu Gly Gly Asp Ala Ser Tyr Gly His Arg Ile Thr Lys Asn Pro Leu
 120 125 130 135
 AGC ACC GCC CAT GCG ACC CTC TGG GGC CCC GCA GAC GCG CTC TAC GAG
 725
 45 Ser Thr Ala His Ala Thr Leu Trp Gly Pro Ala Asp Ala Leu Tyr Glu
 140 145 150

CTG CGC GCC CTC GCA CAC ACC CTC GAC GAG ATC CAC GCA CTG CCG GAG
 773
 Leu Arg Ala Leu Ala His Thr Leu Asp Glu Ile His Ala Leu Pro Glu
 155 160 165

5 GCA GGG AAC CCG CGT CGC AAC GTC ACC CGA TCA ACG GTC GGC CGC AAC
 821
 Ala Gly Asn Pro Arg Arg Asn Val Thr Arg Ser Thr Val Gly Arg Asn
 170 175 180

10 GTC ACC CTG TTC GAC ACC ACC CGC ATG TGG GCA TAC CGG GCC GTC CGG
 869
 Val Thr Leu Phe Asp Thr Thr Arg Met Trp Ala Tyr Arg Ala Val Arg
 185 190 195

15 CAC TCC TGG GGC GGC CCG GTC GCC GAA TGG GAG CAC ACC GTA TTC GAG
 917
 His Ser Trp Gly Gly Pro Val Ala Glu Trp Glu His Thr Val Phe Glu
 200 205 210 215

20 CAC ATC CAC CTA CTG AAC GAG ACG ATC ATC GCC GAC GAA TTC GCC ACA
 965
 His Ile His Leu Leu Asn Glu Thr Ile Ile Ala Asp Glu Phe Ala Thr
 220 225 230

GGC CCC CTC GGC TTG AAC GAA CTT AAG CAC TTA TCT CGA TCC ATT TCC
 1013
 Gly Pro Leu Gly Leu Asn Glu Leu Lys His Leu Ser Arg Ser Ile Ser
 235 240 245

25 CGA TGG GTC TGG CGC AAC TTC ACC CCC GAA ACC TTC CGC GCA CGC CAG
 1061
 Arg Trp Val Trp Arg Asn Phe Thr Pro Glu Thr Phe Arg Ala Arg Gln
 250 255 260

30 AAA GCG ATC AGC CTC CGT GGA GCA TCC AAA GGC GGC AAA GAA GGC GGC
 1109
 Lys Ala Ile Ser Leu Arg Gly Ala Ser Lys Gly Gly Lys Glu Gly Gly
 265 270 275

35 CAC AAA GGC GGC ATT GCC AGT GGC GCA TCA CGG CGC GCC CAT ACC CGT
 1157
 His Lys Gly Gly Ile Ala Ser Gly Ala Ser Arg Arg Ala His Thr Arg
 280 285 290 295

CAA CAG TTC TTG GAG GGT CTC TCA TGACCACACG TGAACGTCTC CCCCGCAACG
 1211
 Gln Gln Phe Leu Glu Gly Leu Ser
 300

40 GCTACAGCAT CGCCGCTGCT GCGAAAAAGC TCGGTGTCTC CGAGTCCACC GTCAAGCGGT
 1271

GGACTTCCGA GCCACGCGAG GAGTTCGTGG CCCGCGTTGC CGCACGCCAC GCGCGGATTG
 1331

45 GTGAGCTCCG CTCGGAGGGT CAGAGCATGC GTGCGATTGC TGCCGAGGTC GGGGTTTCCG
 1391

TGGGCACCGT GCACTACGCG CTGAACAAGA ATCGAACTGA CGCATGACCG TAACGCCGCA
1451

CGATGAGCAT TTTCTTGATC GTGCACCGCT TGGCACTACG TTCGCGTGCG GTTGACACAGT
1511

5 GCGCGCCACG TTCTTATCCT GCGGCCATTG TGGCTACAGC CAATGGGGGG CATCAGCAAC
1571

GGACGTTGAA CCCGGTGGGC AAGTGTTACT CAGGGGGACA TGCCAGTCT GCGGCGCTCG
1631

10 GATTGACGGT ATGGCAGTCG TGCATGCGGC CCCACCGTCA AACTCATTCA GGTATCAGTG
1691

AGAACCCTCA TGGCACCCCC TCGTGACACG TTCTCGTTGC GATCAGCTGC TGTGCGTGCG
1751

GGCGTGAGCG TTTCTACGCT GCGGCGCAGG AAATCAGAGC TTGAGGCTGC CGGAGCGACG
1811

15 GTAGACCCGT CCGGTTGGGT GGTGCCACTG CGTGCACTCA AGGTCGTTTT TGGGGTGTC
1871

GATGAGACCT CGAATGCGCC CGGTCATGAC GCTGAGTTAG TGGCGCAGCT GCGCTCTGAG
1931

20 AACGAGTTTT TACGGCGTCA GGTGAGCAG CAGGCGCGCA CGATCGAACG GCAGGCTGAG
1991

GCACACGCGG TGGTCTCAGC GCAGCTCACA CGGGTTGGCC AGCTTGAGGC CGGCGACGCA
2051

GCAGCACCGA CACTGGCACC CGTTGAAAGG CCGGCTCCGC GACGGCGGTG GTGGCAGCGT
2111

25 CCGTAGCGGT CAGGATCGCT CTGGCGTGAC GAGTGTGTCT GGCAGTGCGA ACAGTTGCTC
2171

GACCAAGTGGC AGCAGAAGCG AGATCGCTGC GTGGTGCTGT TCCTCGGTCA GTTCGTCGAG
2231

30 GACTGGCGGG TCTTGCTGCG TCCAGCCGAT CGCCTCGGCG GCCAAGGTCA GTTCCAAGCT
2291

GTGCCAACGC ACACGCCCT CGGCTGACAG CTGAGTCTCG AACTGTGCAA CTGGACCGGC
2351

CGGAAGATGC ACGTTGCCGA GGTGCTGAGT GGCCAAGCGC ACGTCAAAGA GTGCTGCTTC
2411

35 GTAGCCGCGC AGAAATGGCA GTGCTCGGTC GATTCCGATC GGCCTGCCCA GGTACATTCC
2471

GGGCCGCTTG ATGAACGCCT CCGCGTAGAA GCGCACCGTT CTCGGCCCCG CCTCGTGATC
2531

TGTCACTGTG CACGCTCCTC TCGATGGTTC TCGACGCTAC CGGAGACCAC CGACGTTTCAT

2591

GCCCAGCGCA GCGACCTGAA AGGACCAAGC CGAGTTAGCC GTGCTAACCG TATAGCTTGC
2651

5

TCCGTCGCCT CTGAGGGCAA CCACCTGCGC AGCAGGTGGG CGGCAGCCCC CGCGCAAGCG
2711

CCTACCGGGT TTGGGCACAG CCCATAAATC AACGCCTCCG GTGTTGAAGC GATCGTGTGT
2771

CACGATTGCT ATGCTTGCTA CCCCTTCAGG GTTTTCGTAT ACACAAATCA AGTTTTTTTCG
2831

10

TATACGCTAA TGCCATGAGT GAGCATCTAC TGCACGGCAA GCCCGTCACC AACGAGCAGA
2891

TTCAGGCATG GGCAGACGAG GCCGAGGCCG GATACGACCT GCCCAAATC CCCAAGCCAC
2951

15

GGCGCGGACG CCCGCCCCGTA GGAGACGGTC CGGGCACCGT CGTACCCGTG CGTCTCGACG
3011

CGGCCACCGT TGCCGCTCTC ACAGAACGAG CAACAGCCGA GGGCATCAG AACC GTTCAG
3071

ACGCGATCCG AGCCGCAGTC CACGAGTGA CACGGGTTGC CTGACCTCCA CGACTCAGCA
3131

20

CGCAAGCACT ACCAACGAGA CCGGCTCGAC GACACGGCCG TGCTCTACGC GGCCACCCAC
3191

GTTCTCAACT CCCGGCCACT CGACGACGAA GACGACCCGC GCCGCTGGCT CATGATCGGA
3251

25

ACCGACCCAG CAGGCCGCCT ACTCGAACTC GTCGCACTGA TCTACGACGA CGGCTACGAA
3311

CTGATCATCC ACGCAATGAA AGCCCGCACC CAATACCTCG ACCAGCTCTA ACCAAGAAAG
3371

GAACCTGATG AGCGACCAGC TAGACAGCGA CCGCAACTAC GACCCGATGA TCTTCGACGT
3431

30

GATGCGCGAG ACCGCGAACC GCGTCGTGCG CACGTACGTT GCATGGGAAG ATGAAGCCGC
3491

TGATCCCCGC GAGGCTGCGC ACTGGCAGGC CGAGCGATTG CGCACCCGGC ACGAGGTGCG
3551

35

CGCC
3555

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 303 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

5 Met Asp Ser Phe Glu Thr Leu Phe Pro Glu Ser Trp Leu Pro Arg Lys
 1 5 10 15
 Pro Leu Ala Ser Ala Glu Lys Ser Gly Ala Tyr Arg His Val Thr Arg
 20 25 30
 10 Gln Arg Ala Leu Glu Leu Pro Tyr Ile Glu Ala Asn Pro Leu Val Met
 35 40 45
 Gln Ser Leu Val Ile Thr Asp Arg Asp Ala Ser Asp Ala Asp Trp Ala
 50 55 60
 Ala Asp Leu Ala Gly Leu Pro Ser Pro Ser Tyr Val Ser Met Asn Arg
 65 70 75 80
 15 Val Thr Thr Thr Gly His Ile Val Tyr Ala Leu Lys Asn Pro Val Cys
 85 90 95
 Leu Thr Asp Ala Ala Arg Arg Arg Pro Ile Asn Leu Leu Ala Arg Val
 100 105 110
 20 Glu Gln Gly Leu Cys Asp Val Leu Gly Gly Asp Ala Ser Tyr Gly His
 115 120 125
 Arg Ile Thr Lys Asn Pro Leu Ser Thr Ala His Ala Thr Leu Trp Gly
 130 135 140
 Pro Ala Asp Ala Leu Tyr Glu Leu Arg Ala Leu Ala His Thr Leu Asp
 145 150 155 160
 25 Glu Ile His Ala Leu Pro Glu Ala Gly Asn Pro Arg Arg Asn Val Thr
 165 170 175
 Arg Ser Thr Val Gly Arg Asn Val Thr Leu Phe Asp Thr Thr Arg Met
 180 185 190
 30 Trp Ala Tyr Arg Ala Val Arg His Ser Trp Gly Gly Pro Val Ala Glu
 195 200 205
 Trp Glu His Thr Val Phe Glu His Ile His Leu Leu Asn Glu Thr Ile
 210 215 220
 Ile Ala Asp Glu Phe Ala Thr Gly Pro Leu Gly Leu Asn Glu Leu Lys
 225 230 235 240
 35 His Leu Ser Arg Ser Ile Ser Arg Trp Val Trp Arg Asn Phe Thr Pro
 245 250 255
 Glu Thr Phe Arg Ala Arg Gln Lys Ala Ile Ser Leu Arg Gly Ala Ser
 260 265 270

Lys Gly Gly Lys Glu Gly Gly His Lys Gly Gly Ile Ala Ser Gly Ala
 275 280 285

Ser Arg Arg Ala His Thr Arg Gln Gln Phe Leu Glu Gly Leu Ser
 290 295 300

5 (2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 85 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Thr Thr Arg Glu Arg Leu Pro Arg Asn Gly Tyr Ser Ile Ala Ala
 1 5 10 15

Ala Ala Lys Lys Leu Gly Val Ser Glu Ser Thr Val Lys Arg Trp Thr
 15 20 25 30

Ser Glu Pro Arg Glu Glu Phe Val Ala Arg Val Ala Ala Arg His Ala
 35 40 45

Arg Ile Arg Glu Leu Arg Ser Glu Gly Gln Ser Met Arg Ala Ile Ala
 50 55 60

Ala Glu Val Gly Val Ser Val Gly Thr Val His Tyr Ala Leu Asn Lys
 20 65 70 75 80

Asn Arg Thr Asp Ala
 85

(2) INFORMATION FOR SEQ ID NO: 4:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 59 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATTCAGCT TGTGACGTT AACCTGCAGG CATGCGGATC CGGTACCGAT ATCAGATCT
 59

(2) INFORMATION FOR SEQ ID NO: 5:

- 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCGAAGATCT GATATCGGTA CCGGATCCGC ATGCCTGCAG GTTAACGTCG ACAAGCTTG
59

(2) INFORMATION FOR SEQ ID NO: 6:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GTACCGGCCG CTGCGGCCAA GCTT
24

(2) INFORMATION FOR SEQ ID NO: 7:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GATCAAGCTT GGCCGCAGCG GCCG
24

(2) INFORMATION FOR SEQ ID NO: 8:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

AAACTGCAGC TGCTGGCTTG CGCCCGATGC TAGTC
35

(2) INFORMATION FOR SEQ ID NO: 9:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 76 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AAACTGCAGC AGCTGGGCAG GCCGCTGGAC GGCCTGCCCT CGAGCTCGTC TAGAATGTGC
60

15 TGCCGATCCT GGTTC
76

(2) INFORMATION FOR SEQ ID NO: 10:

- 20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

25 CTAGTCTAGA CACCGATGAG GAAACCCGAT GA
32

(2) INFORMATION FOR SEQ ID NO: 11:

- 30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCAAGCTTC TCGAGTCAGT GGTCGCTGGG CGCGCG

36

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

10 GGAGATCTAGATCGATATCTCGAG
24

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

20 GATCCTCGAGATATCGATCTAGATCTCCGC
30